STATISTICS IN MEDICINE
*Statist. Med.* 2005; **24**:2659–2668
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2135

# Using the Peters–Belson method to measure health care disparities from complex survey data¶

B. I. Graubard[1,*,†], R. Sowmya Rao[2,‡] and Joseph L. Gastwirth[3,§]

[1] *Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Executive Plaza South, Room 8024, 6120 Executive Boulevard, MSC 7244, Bethesda, MD 20892, U.S.A.*
[2] *Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138, U.S.A.*
[3] *Department of Statistics, 315 Funger Hall, George Washington University, Washington, DC 20052, U.S.A.*

## SUMMARY

The Peters–Belson (PB) method uses regression to assess wage discrimination and can also be used to analyse disparities for a variety of health care issues, e.g. cancer screening. The PB method estimates the proportion of an overall disparity that is not explained by the covariates in the regression, e.g. education, which may be due to discrimination. This method first fits a regression model with individual-level covariates to the majority/advantaged group and then uses the fitted model to estimate the expected values for minority-group members had they been members of the majority group. The data on disparities in health care available to biomedical researchers differ from data used in legal cases as it is often obtained from large-scale studies or surveys with complex sample designs involving stratified multi-stage cluster sampling. Sample surveys with a large representative sample of various racial/ethnic groups and the extensive collection of important social–demographic variables provide excellent sources of data for assessing disparity for a wide range of health behaviours. We extend the PB method for multiple logistic and linear regressions of simple random samples to weighted data from complex designed survey samples. Because of the weighting and complex sample designs, we show how to apply the Taylor linearization method and delete-one-group jackknife methods to obtain estimates of standard errors for the estimated disparity. Data from the 1998 National Health Interview Survey on racial differences in cancer screening among women is used to illustrate the PB method. Published in 2005 by John Wiley & Sons, Ltd.

KEY WORDS: survey methods; Taylor linearization variance; delete-one-group jackknife variance; logistic regression; multiple linear regression

*Correspondence to: B. I. Graubard, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Executive Plaza South, Room 8024, 6120 Executive Boulevard, MSC 7244, Bethesda, MD 20892, U.S.A.
†E-mail: graubarb@mail.nih.gov
‡E-mail: sowmya_rao@abtassoc.com
§E-mail: jlgast@gwu.edu


Contract/grant sponsor: National Science Foundation; contract/grant number: SES-0317956

Published in 2005 by John Wiley & Sons, Ltd.

*Received 12 March 2004*
*Accepted 4 November 2004*

## 1. INTRODUCTION

The Peters–Belson (PB) approach (called the Blinder–Oaxaca approach in the economics literature) has been used in wage discrimination studies [1–3] and race (sex) discrimination cases [4, 5] to predict the wage a minority (female) individual would have had if they were a white (male) with the same qualifications (see Reference [6] for a review of methods for estimating race and gender disparity in the labour market). The conventional regression approach includes a dummy variable to identify sex or race/ethnicity, implicitly assuming a common amount (degree) of disparity for all minority group members. In contrast, the PB approach produces estimates of the disparity for each minority group member by incorporating their individual characteristics. This feature has made the PB approach useful in assessing possible discrimination in salary or hiring practices as the individual estimates can determine what each individual deserves. The statistical properties of the estimated disparity and associated tests and confidence intervals have been obtained by Gastwirth and Greenhouse [7] and Nayak and Gastwirth [8] when the data comes from a simple random sample.

National surveys are excellent sources of data for producing population-based estimates of disparities for various outcomes and various disadvantaged groups. For example, the 1998 National Health Interview Survey (NHIS) collects a variety of measures of health behaviour and of socio-demographic variables, which have been used to assess disparity of minority racial/ethnic groups in being screened for cancer [9]. However, these types of surveys have complex sampling designs that involve cluster and stratified sampling and sample weighting, which need to be accounted for in the analysis to obtain approximately unbiased parameter estimates of the population with appropriate standard errors. This paper develops appropriate methodology for use of multiple logistic or linear regression in the PB framework applied to complex survey data. Special attention will be paid to using the sample weights in the estimation of disparity and to account for the weighting, stratification, and clustering of the sampling in the estimation of the standard errors of the disparity measure, in which we use a design-based approach for estimating the disparity and its variance; other variance estimation methods that rely more heavily on the correctness of the regression model are available (see Reference [10, pp. 310–314, 323–326]).

Two popular methods for obtaining standard errors of estimators of complex survey data where the estimators are non-linear differentiable functions are: (i) delete-one-group jackknife [11, pp. 29–31]; and (ii) Taylor linearization [11, pp. 23–28] [12, 13]. The delete-one-group jackknife method is a replication method that is based on repeated estimation of the parameter after removing a subset of the data as determined by the clusters of the units used in the sampling. The delete-one-group jackknife was applied by Rao *et al.* [9] for a PB analysis of the 1998 NHIS that used logistic regression. The Taylor linearization method uses the delta method to approximate a non-linear estimator by an estimator of a total. Standard analytical formulas are then used to estimate the standard error of this total. The jackknife and Taylor linearization methods are asymptotically consistent for complex sampling [14] and are shown to yield similar results for the 1998 NHIS.

This paper is laid out as follows: Section 2 describes the PB method for estimating disparity from weighted survey data. The description of the Taylor linearization and the delete-one-group jackknife methods for estimating standard errors of disparity with data from surveys with stratified multistage cluster samples is presented in Section 3. The PB approach is illustrated in Section 4 for estimating racial/ethnic disparity in cancer screening among women in the 1998

NHIS. The paper concludes with a Discussion, and Appendix A provides details for estimating standard errors of disparity using the PB method that is based on logistic regression.

## 2. THE PETERS–BELSON APPROACH USING LOGISTIC REGRESSION

From a survey of $n$ subjects, suppose we are interested in evaluating the disparity in a binary outcome $(Y)$, e.g. having a recent screen for colorectal cancer, between two groups of the subjects, $R_0$ and $R_1$, e.g. white $(R_0)$ and black $(R_1)$ women. $R_0$ denotes the reference group, advantaged/majority group, and $R_1$ denotes the disadvantaged/minority group that is being compared to $R_0$. Each sampled person $i$ in the survey has a 0–1 binary outcome $y_i$, a $p \times 1$ covariate vector $x_i$, a sample weight $w_i$, which is the inverse of the probability of including the person in the sample, and an indicator variable denoting group membership that is given by $\delta_{R_k i} = 1$ if the person $i$ is from group $R_k$, and $\delta_{R_k i} = 0$, otherwise, $k = 0$, 1. The (sample) weighted estimate of the crude proportion of individuals from group $R_k$ with outcome $y = 1$ is

$$p_{R_k} = \frac{\sum_{j=1}^{n} w_j \delta_{R_k j} y_j}{\sum_{j=1}^{n} w_j \delta_{R_k j}}$$

We fit a logistic regression to the observations from the reference group $R_0$ with $\Pr(y_i = 1 | x_i, i \in R_0) = \exp(x_i' \beta_{R_0}) / [1 + \exp(x_i' \beta_{R_0})]$, where the $x_i'$ is the transpose of $x_i$ and $\beta_{R_0}$ is the $p \times 1$ vector of regression coefficients. The (sample) weighted pseudo-likelihood

$$\prod_{i=1}^{n} \left[ \frac{\exp(x_i' \beta_{R_0})}{1 + \exp(x_i' \beta_{R_0})} \right]^{\delta_{R_0 i} w_i y_i} \left[ \frac{1}{1 + \exp(x_i' \beta_{R_0})} \right]^{\delta_{R_0 i} w_i (1 - y_i)} \tag{1}$$

is maximized to obtain $\hat{\beta}_{R_0}$, the (sample) design consistent estimate of $\beta_{R_0}$ [11, p. 101]. The PB method assumes that the probability of a positive outcome for a person $i$ in group $R_1$ should follow the same logistic model as a person from the reference group $R_0$, so each person $i$ in group $R_1$ has a predicted outcome

$$\hat{p}_{R_0 i} = \frac{\exp(x_i' \hat{\beta}_{R_0})}{1 + \exp(x_i' \hat{\beta}_{R_0})} \tag{2}$$

Thus, using the logistic model for group $R_0$, the estimated expected proportion of individuals with $y = 1$ that are in group $R_1$ is

$$\hat{p}_{R_0, R_1} = \frac{\sum_{j=1}^{n} w_j \delta_{R_1 j} \hat{p}_{R_0 j}}{\sum_{j=1}^{n} w_j \delta_{R_1 j}}$$

The estimated difference between the crude proportions of individuals in $R_0$ and $R_1$ with $y = 1$, $p_{R_0} - p_{R_1}$, which is referred to as the 'overall disparity', can be partitioned as

$$p_{R_0} - p_{R_1} = (p_{R_0} - \hat{p}_{R_0, R_1}) + (\hat{p}_{R_0, R_1} - p_{R_1}) \tag{3}$$

The difference $p_{R_0} - \hat{p}_{R_0,R_1}$ estimates that part of the overall disparity that can be explained by the covariates and the difference $\hat{p}_{R_0,R_1} - p_{R_1}$ estimates the 'unexplained disparity', which is that part of the overall disparity that *cannot* be explained by the covariates $x$ in the logistic model. The proportion of the overall disparity that is not explained by the covariates may be due to discrimination or to omitted relevant covariates that are distributed differently between the groups. The proportion or percent of disparity that is explained by the covariates is estimated by $D_{R_0}/D$ or $100 \times D_{R_0}/D$, which is referred to as 'per cent explained', where $D = p_{R_0} - p_{R_1}$ and $D_{R_0} = p_{R_0} - \hat{p}_{R_0,R_1}$.

## 3. VARIANCE ESTIMATION FOR DISPARITY

### *3.1. Taylor linearization*

The estimator of the PB measure of disparity of the 'per cent explained' is a complex non-linear function of the survey data involving the predicted outcomes from the logistic regression model and the sample weights. In addition, the estimate of disparity utilizes observations that can be dependent because of cluster sampling of the survey. Because the estimator of the 'per cent explained' is a differentiable function of the data, the Taylor linearization method for variance can be used to estimate its standard error. In general, this variance estimation method involves first computing the Taylor deviate of the estimator for each observation ($z_i$), which is a measure of the influence or change in the value of the statistic when that observation is deleted. Shah [12, 13] showed that the Taylor deviate is easily obtained by differentiating a sample-weighted estimator with respect to its weights (for similar results see also References [15–17]). An expression for the Taylor deviate $z_i$ of the estimator of the 'per cent explained' based on logistic regression modelling is derived in Appendix A. Next, because the sample-weighted sum $\sum_{i=1}^{n} w_i z_i$ is an asymptotically consistent linear approximation to the complex estimator [15], we can use standard survey sampling variance formulas for estimating the variance of the weighted sum [18] to estimate the variance of the complex estimator.

Standard software is available for estimating variances for totals such as $\sum_{i=1}^{n} w_i z_i$ under a variety of complex sample designs (e.g. Reference [19]), including household surveys such as the NHIS with multistage stratified cluster sampling. For surveys such as the NHIS, the target population of individuals is partitioned into first stage clusters called primary sampling units (PSUs), which are usually counties or cities. The PSUs are grouped into $H$ strata that are formed to be approximately homogeneous with respect to specific demographic characteristics of the PSU populations, e.g. population size. At the first stage of sampling $t_h$ PSUs are randomly sampled from each stratum $h = 1, \ldots, H$. At the second and further stages, stratification and cluster sampling can be used to sample individuals within the sample PSUs. Let $t_{hi},\ i = 1, \ldots, t_h$ be the number of individuals sampled from sampled PSU $i$ in stratum $h$. For these later stages of sampling in 1998 NHIS, clusters of households with high concentrations of Hispanic and black populations are oversampled to increase the sample sizes of individuals from these groups. (For further details about the sample design of the 1998 NHIS see Reference [20].) Often the proportion of PSUs sampled at the first stage from each of the strata is small so that the sample of the PSUs can be approximated as a stratified with-replacement sample. In this case for a multistage stratified cluster sample the variance estimator of the weighted total of the Taylor deviates simplifies and is given by

$$\hat{\mathrm{var}}\left(\sum_{h=1}^{H}\sum_{i=1}^{t_h}\sum_{j=1}^{t_{hi}} w_{hij}z_{hij}\right) = \sum_{h=1}^{H}\frac{t_h}{t_h-1}\sum_{j=1}^{t_h}(z_{hj}-\bar{z}_h)^2 \tag{4}$$

where $z_{hij}$ and $w_{hij}$, respectively, are the Taylor deviate and sample weight for sampled individual $j$ from sampled PSU $i$ from stratum $h$, $z_{hi}=\sum_{j=1}^{t_{hi}} w_{hij}z_{hij}$, and $\bar{z}_h = 1/t_h \sum_{i=1}^{t_h} z_{hi}$.

### 3.2. Delete-one-group jackknife

When the estimator is a differentiable function of the data as is the case for the disparity estimator, we can use a delete-one-group jackknife method for estimating the variance of the estimator [11, pp. 29–31]. The jackknife estimator will depend on the sample design. For the NHIS multistage stratified cluster sample design the jackknife leaves all observations from one PSU out at a time and recomputes the estimator each time. We first compute the overall estimate of disparity $\hat{\theta}=D_{R_0}/D$. Then we re-compute the disparity $\hat{\theta}_{(hi)}$ leaving out observations from PSU $i$ in stratum $h$ and increasing the sample weights of the retained obervations in stratum $h$ by the factor $t_h/(t_h-1)$. The jackknife variance estimator is computed as:

$$\sum_{h=1}^{H}\frac{t_h-1}{t_h}\sum_{i=1}^{t_h}(\hat{\theta}_{(hi)}-\hat{\theta})^2 \tag{5}$$

## 4. AN APPLICATION

The data for this application is taken from the 1998 NHIS. The NHIS is a nationally representative household interview survey of the civilian non-institutionalized population of the U.S. The sample design of the 1998 NHIS is complex with stratified multistage probability cluster sampling where at the first stage of sampling 2 PSUs in each of the 339 strata were selected [20]. In this survey, the U.S. Department of Health and Human Services sponsored a Health Prevention Supplement (HPS) to the core of the NHIS, which included questions on the utilization of cancer screening. In each of the eligible 38 209 households selected for interview, an adult respondent ($\geqslant$18 years old) was randomly selected to complete the HPS.

Following Rao *et al.* [9], we estimated disparities in cancer screening between white, black, and Hispanic women for: (i) colorectal cancer exam—fecal occult blood test during the past 2 years or endoscopy during the past 3 years; (ii) digital rectal exam during the past 2 years; (iii) Pap test during the past 3 years; and (iv) mammogram during the past 2 years. Due to age-specific guidelines for screening use, and education, which is presumed to be essentially completed by age 25, the analysis was restricted to subjects over the age of 49 years for colorectal cancer screening and digital rectal exam, over the age of 39 years for mammography, and over the age of 24 for Pap test [21, 22]. For each type of screening test logistic regression models were fit to the data for the majority group (usually white). The covariates predictive of mammography use were age, age$^2$, income (below the poverty level [poor], 100–199 per cent of poverty level [near poor], and $\geqslant$200 per cent of poverty level [middle/high]), education (less than high school graduate, high school graduate, and at least some college), region of residence (Northeast, West, Midwest, and South), metropolitan

Table I. Disparity in recent cancer screening* among women by race/ethnicity.

| Type of screening | Race | Sample size | Observed (per cent) | Predicted† (per cent) | Per cent explained‡ | Standard error of per cent explained (jackknife) | Standard error of per cent explained (Taylor linearization) |
|---|---|---|---|---|---|---|---|
| Colorectal | White | 4016 | 31.1 | | | | |
| (Age 50+) | Black | 691 | 24.7 | 26.5 | 72.4 | 26.1 | 26.0 |
| | Hispanic | 526 | 19.2 | 26.5 | 39.1 | 10.3 | 10.3 |
| Digital rectal exam | White | 3993 | 43.4 | | | | |
| (Age 50+) | Black | 683 | 38.2 | 38.5 | 93.4 | 47.6 | 47.2 |
| | Hispanic | 518 | 32.3 | 35.0 | 75.7 | 21.5 | 21.5 |
| Mammography | White | 5888 | 66.7 | | | | |
| (Age 40+) | Black | 1049 | 63.9 | 60.6 | 224.9 | 156.7 | 152.4 |
| | Hispanic | 928 | 60.1 | 56.6 | 154.3 | 47.7 | 47.5 |
| Pap test | Black | 2354 | 80.1 | | | | |
| (Age 25+) | White | 11 077 | 77.0 | 77.6 | 79.2§ | 26.6 | 26.5 |
| | Hispanic | 2478 | 73.8 | 80.0 | 1.8§ | 8.8 | 8.8 |

*For colorectal screening, 'recent' is if the respondent reported fecal occult blood test for screening during the past 2 years or endoscopy for screening during the past 3 years; for digital rectal exam and mammography, 'recent' is defined as during the past 2 years preceding the interview, and for Pap test, 'recent' defined as during the 3 years preceding the interview.

†Predicted using the logistic regression model fit among the majority group (white) except for Pap test where the majority group was black.

‡((Observed (white) – predicted (black/Hispanic))/(observed (white) – observed (black/Hispanic)))∗100.

§((Observed (black) – predicted (white/Hispanic))/(observed (black) – observed (white/Hispanic)))∗100.

statistical area (in MSA or not), health insurance coverage (yes or no), and usual source of care (yes or no). As education, MSA and health insurance coverage were not predictive of colorectal screening they were not included. MSA and region of residence were not predictive of digital rectal exam and were excluded from that model. Only age and usual source of care were predictive of Pap test.

The group with the highest rate of screening ('best' group) was used as the reference group ($R_0$). This meant that white women were the reference group for all cancer screening modalities except for Pap test screening where black women were the reference group with the highest screening rate.

All estimates were weighted using the NHIS sample weights. We computed variances for the estimates using the Taylor linearization approach (4) and compared them to the stratified cluster sample version of the delete-one-group jackknife variance estimator (5). The procedure IML in the statistical software package of SAS V8.2 [23] was used to compute the Taylor deviates, and SUDAAN V8.0 [19] was used to compute the variances of the weighted sum of the deviates.

The racial/ethnicity disparities in cancer screening rates for women are displayed in Table I. The covariates that explain black–white differences do not explain as much of the Hispanic–white differences and Hispanic–black difference for Pap tests. The standard errors obtained using the jackknife approach, as well as, the Taylor linearization approach are presented in the last two columns of Table I. The standard errors obtained by both these methods closely agree. Even though the sample sizes in the NHIS were moderately large, we found that the 'per cent explained' had relatively large standard errors, resulting in fairly wide confidence intervals. This is because the estimated crude differences in screening rates between any two groups appears in the denominator of the 'per cent explained', and these differences can be small resulting in high variability of the 'per cent explained'.

## 5. DISCUSSION

In this paper we describe the PB method for estimating disparity using logistic regression for the analysis of complex survey data. We describe the estimation of standard errors using both the Taylor linearization approximation of the estimated 'per cent explained' and the delete-one-group jackknife method. We illustrate the PB method for survey with an application to the NHIS in evaluating disparities in cancer screening rates by ethnicity among women.

Large national surveys (e.g. 1998 NHIS) are excellent data sources for studying differences in health practices and behaviour between the major and minority groups because they often over-sample African-Americans and Hispanics. Thus, the sample sizes are adequate to accurately estimate between-group differences in the predictors of outcomes such as cancer screening for African-Americans, Hispanics and whites. In addition, the sample weighted estimation provides U.S.-based population estimates of disparity.

Originally Peters [24] and Belson [25] developed their measure of disparity for a multiple linear regression model fitted to observations on a dependent variable $y$ and covariates $x$ from the reference group $R_0$. For survey data, the estimated linear regression coefficients, $\hat{\beta}_{R_0}$, can be obtained using weighted least squares with the weights being the sample weights [11, pp. 92–93]. The disparity measure, $D_{R_0}/D \times 100$ per cent, is estimated by letting the $\hat{p}_{R_0 j} = x_j' \hat{\beta}_{R_0}$ for members $j$ from the minority group $R_0$. The variance estimator for this disparity measure can be obtained for a multistage stratified cluster sample by using the general expression for the variance estimator given in equation (4) with the Taylor deviates for the logistic regression case modified for multiple linear regression by letting $\partial \hat{p}_{R_0 j}/\partial \hat{\beta}_{R_0} = x_j$ and $\partial \hat{\beta}_{R_0}/\partial w_i = [\sum_{j=1}^{n} w_j \delta_{R_0 j} x_j x_j']^{-1} \delta_{R_0 i} x_i (y_i - \hat{p}_{R_0 i})$ in equation (A2) in Appendix A. The expression for $\partial \hat{\beta}_{R_0}/\partial w_i$ was obtained from differentiation of the weighted least squares estimating equations.

It has been noted that when the covariate distributions between the groups do not overlap well the PB method can yield a biased estimate of the disparity because the regression model that has been fitted to the majority group may not be appropriate for the region of the covariate space covered only by the distribution of the minority group [26]. Non-parametric approaches have been proposed that may yield a less biased estimate when there is little overlap in the covariate distributions [26, 27]. In our example, the covariates were categorical, except for age, and there was reasonable overlap between the covariate distributions.

The NHIS sample weights are random variables because they include adjustments for post-stratification and non-response to the inverse of the probabilities of selection. However, as pointed out by a reviewer, our method of variance estimation does not account for variability in these weights, as they are treated as fixed quantities. Canty and Davison [28] empirically show that ignoring postratification can result in under-estimation of this variance. Korn and

Graubard [11, pp. 35–38] however, argue that poststratification often tends to reduce variability. In practice many national surveys like the NHIS do not provide sufficient information about the weighting to enable analysts to account for these adjustments in the variance estimation process. There are national surveys (e.g. third National Health and Nutrition Examination Survey, see Reference [11, p. 35]) that provide replicate weights that can be used to estimate variances, e.g. from balanced-half sample repeated replication, where the variability from the adjustments to the sample weights are taken into account in the variance estimation [29].

The approach that we used to derive the Taylor linearization is quite general and can be applied to other complex statistics such as estimating the variances for the regression coefficients from Cox proportional hazard regression of a complex survey [12, 13].

In the NHIS example, we found that the standard errors obtained using the Taylor linearization method were close to those obtained using the jackknife method. Although the jackknife method is easy to apply, it is very computer intensive, especially for a survey with a large number of PSUs. The Taylor linearization method is less computer intensive and can be easily implemented by computing a Taylor deviate for each observation and then use a software package (e.g. SUDAAN) to estimate the variances of the sample weighted totals of those deviates.

## APPENDIX A: DERIVATION OF THE TAYLOR DEVIATE FOR THE PETERS–BELSON MEASURE OF DISPARITY UNDER LOGISTIC REGRESSION

Using the approach of Shah [12, 13], the Taylor deviates ($z_i$) for the estimated disparity $D_{R_0}/D$ are:

$$z_i = \frac{\partial D_{R_0}/D}{\partial w_i} = \frac{\partial(1 - D_{R_1}/D)}{\partial w_i} = \frac{-1}{D}\frac{\partial D_{R_1}}{\partial w_i} + \frac{D_{R_1}}{D^2}\frac{\partial D}{\partial w_i} \tag{A1}$$

where $D_{R_1} = \hat{p}_{R_0,R_1} - p_{R_1}$. We proceed to obtain an expression for the $z_i$ by computing the expressions for $\partial D_{R_1}/\partial w_i$, $\partial D/\partial w_i$ and substituting them into (A1).

$$\text{Since } D_{R_1} = \frac{\sum_{j=1}^n w_j \delta_{R_1 j}\, \hat{p}_{R_0 j} - \sum_{j=1}^n w_j \delta_{R_1 j}\, y_j}{\sum_{j=1}^n w_j \delta_{R_1 j}}$$

$$\frac{\partial D_{R_1}}{\partial w_i} = \frac{-1}{\sum_{j=1}^n w_j \delta_{R_1 j}}\left[\delta_{R_1 i}(y_i - \hat{p}_{R_0 i}) - \left(\sum_{j=1}^n w_j \delta_{R_1 j}\frac{\partial \hat{p}_{R_0 j}}{\partial \hat{\beta}_{R_0}}\right)'\frac{\partial \hat{\beta}_{R_0}}{\partial w_i}\right] - \frac{\delta_{R_1 j}}{\left(\sum_{j=1}^n w_j \delta_{R_1 j}\right)^2}\sum_{j=1}^n w_j \delta_{R_1 j}(y_j - \hat{p}_{R_0 j}) \tag{A2}$$

where chain rule is used to obtain the second expression in the square brackets. From equation (2) in the main text, we obtain the $p \times 1$ vector $\partial \hat{p}_{R_0 j}/\partial \hat{\beta}_{R_0} = \hat{p}_{R_0 j}(1 - \hat{p}_{R_0 j})x_j$. An expression for $\partial \hat{\beta}_{R_0}/\partial w_i$ is obtained from differentiating the weighted pseudo-likelihood estimating equations for $\beta_{R_0}$ from the logistic regression model [11, p. 101], with respect to the sample weights:

The weighted pseudo-likelihood estimating equations for $\beta_{R_0}$ evaluated at $\hat{\beta}_{R_0}$ are

$$\sum_{j=1}^{n} w_j \delta_{R_0 j} x_j (y_j - \hat{p}_{R_0 j}) = 0$$

Differentiating these estimating equations with respect to $w_i$

$$\delta_{R_0 i} x_i (y_i - \hat{p}_{R_0 i}) - \sum_{j=1}^{n} w_j \delta_{R_0 j} x_j \frac{\partial \hat{p}_{R_0 j}}{\partial \hat{\beta}_{R_0}} \frac{\partial \hat{\beta}_{R_0}}{\partial w_i} = 0$$

and

$$\delta_{R_0 i} x_i (y_i - \hat{p}_{R_0 i}) - \sum_{j=1}^{n} w_j \delta_{R_0 j} x_j x_j' \hat{p}_{R_0 j} (1 - \hat{p}_{R_0 j}) \frac{\partial \hat{\beta}_{R_0}}{\partial w_i} = 0$$

because $\partial \hat{p}_{R_0 j} / \partial \hat{\beta}_{R_0} = \hat{p}_{R_0 j} (1 - \hat{p}_{R_0 j}) x_j$. Solving for $\partial \hat{\beta}_{R_0} / \partial w_i$ produces the desired result.

$$\frac{\partial \hat{\beta}_{R_0}}{\partial w_i} = \left[ \sum_{j=1}^{n} w_j \delta_{R_0 j} x_j x_j' \hat{p}_{R_0 j} (1 - \hat{p}_{R_0 j}) \right]^{-1} \delta_{R_0 i} x_i (y_i - \hat{p}_{R_0 i})$$

Recalling that

$$D = p_{R_0} - p_{R_1} = \frac{\sum_{j=1}^{n} w_j \delta_{R_0 j} y_j}{\sum_{j=1}^{n} w_j \delta_{R_0 j}} - \frac{\sum_{j=1}^{n} w_j \delta_{R_1 j} y_j}{\sum_{j=1}^{n} w_j \delta_{R_1 j}}$$

then

$$\frac{\partial D}{\partial w_i} = \frac{1}{\sum_{j=1}^{n} w_j \delta_{R_0 j}} [\delta_{R_0 i} (y_i - p_{R_0})] - \frac{1}{\sum_{j=1}^{n} w_j \delta_{R_1 j}} [\delta_{R_1 i} (y_i - p_{R_1})] \tag{A3}$$

We obtain the expression for the Taylor deviates $z_i$ by substituting (A2) and (A3) into (A1).

ACKNOWLEDGEMENTS

The research of Professor Gastwirth was supported in part by Grant #SES-0317956 from the National Science Foundation.

REFERENCES

1. Blinder AS. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 1973; **8**:436–455.
2. Oaxaca R. Male–female differentials in urban labor markets. *International Economic Review* 1973; **14**:693–709.
3. Goldin C. *Understanding the Gender Gap*: *An Economic History of American Women*. Oxford University Press: New York, NY, 1990.
4. Gastwirth JL. A clarification of some statistical issues in Watson v. Fort Worth Bank and Trust. *Jurimetrics Journal* 1989; **29**:267–284.
5. Gray MW. Can statistics tell us what we do not want to hear—the case of complex salary structures. *Statistical Science* 1993; **8**:144–158.
6. Altonji JG, Blank RM. Race and gender in the labor market. In *Handbook of Labor Economics*, vol. 3, Ashenfelter O, Card D (eds). Elsevier Science B.V.: New York, 1999; 3143–3259.

7. Gastwirth JL, Greenhouse SW. Biostatistical concepts and methods in the legal setting. *Statistics in Medicine* 1995; **14**:1641–1653.
8. Nayak TK, Gastwirth JL. The Peters–Belson approach to measures of economic and legal discrimination. In *Advances in the Theory and Practice of Statistics*: *A Volume in Honor of Samuel Kotz*, Johnson NL, Balakrishnan N (eds). Wiley: New York, NY, 1997; 587–601.
9. Rao RS, Graubard BI, Breen N, Gastwirth JL. Understanding the factors underlying disparities in cancer screening rates using the Peters–Belson approach: results from the 1998 National Health Interview Survey. *Medical Care* 2004; **42**:789–800.
10. Valliant R, Dorfman AH, Royal RM. *Finite Population Sampling and Inference*: *A Prediction Approach*. Wiley: New York, NY, 2000.
11. Korn EL, Graubard BI. *Analysis of Health Surveys*. Wiley: New York, NY, 1999.
12. Shah BV. Calculus of Taylor deviations. Presented at the Joint Statistical Meetings, 2002.
13. Shah BV. Comment on 'linearization variance estimators for survey data' by A Demnati and JNK Rao. *Survey Methodology* 2004; **30**:29.
14. Krewski D, Rao JNK. Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics* 1981; **9**:1010–1019.
15. Deville J-C. Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology* 1999; **25**:193–203.
16. Demnati A, Rao JNK. Linearization variance estimators for survey data. *Methodology Branch Working Paper*. SSMD-2001-010E, Statistics, Canada, 2001.
17. Demnati A, Rao JNK. Linearization variance estimators for survey data. *Survey Methodology* 2004; **30**:17–26.
18. Cochran WG. *Sampling Techniques*. Wiley: New York, NY, 1977.
19. Shah BV, Barnwell BG, Bieler GS. *SUDAAN User's Manual*, *release 8.0.* Research Triangle Institute: Research Triangle Park, NC, 1997.
20. Botman SL, Moore TF, Moriarity CL *et al.* Design and estimation for the National Health Interview Survey, 1995–2004. *Vital Health Statistics*. National Center for Health Statistics, vol. 2(130), 2000.
21. Breen N, Wagener DK, Brown ML *et al.* Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998. National Health Interview Surveys. *Journal of the National Cancer Institute* 2001; **93**(22):1704–1713.
22. United States Preventive Services Task Force: http://www.ahcpr.gov/clinic/uspstfix.htm. Accessed 15 May 2003.
23. SAS Institute Inc. *SAS/STAT User's Guide*, *version 8.* Cary: NC, 1999.
24. Peters CC. A method of matching groups for experiments with no loss of populations. *Journal of Educational Research* 1941; **34**:606–612.
25. Belson WA. A technique for studying the effects of a television broadcast. *Applied Statistics* 1956; **5**:195–202.
26. Barsky R, Bound J, Charles KK, Lupton JP. Accounting for the black–white wealth gap: a nonparametric approach. *Journal of the American Statistical Association* 2002; **97**:663–673.
27. Heckman JJ, Lochner L, Todd P. Fifty years of Mincer earnings regressions. *Working Paper 9732*, Nation Bureau of Economic Research, 2003.
28. Canty AJ, Davison AC. Resampling-based variance estimation for labour force surveys. *The Statistician* 1999; **48**:379–391.
29. Rust K, Rao JNK. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* 1996; **5**:283–310.